Communication Theory II

Lecture 6: Review on Probability Theory (cont’d)

Ahmed Elnakib, PhD

Assistant Professor, Mansoura University, Egypt

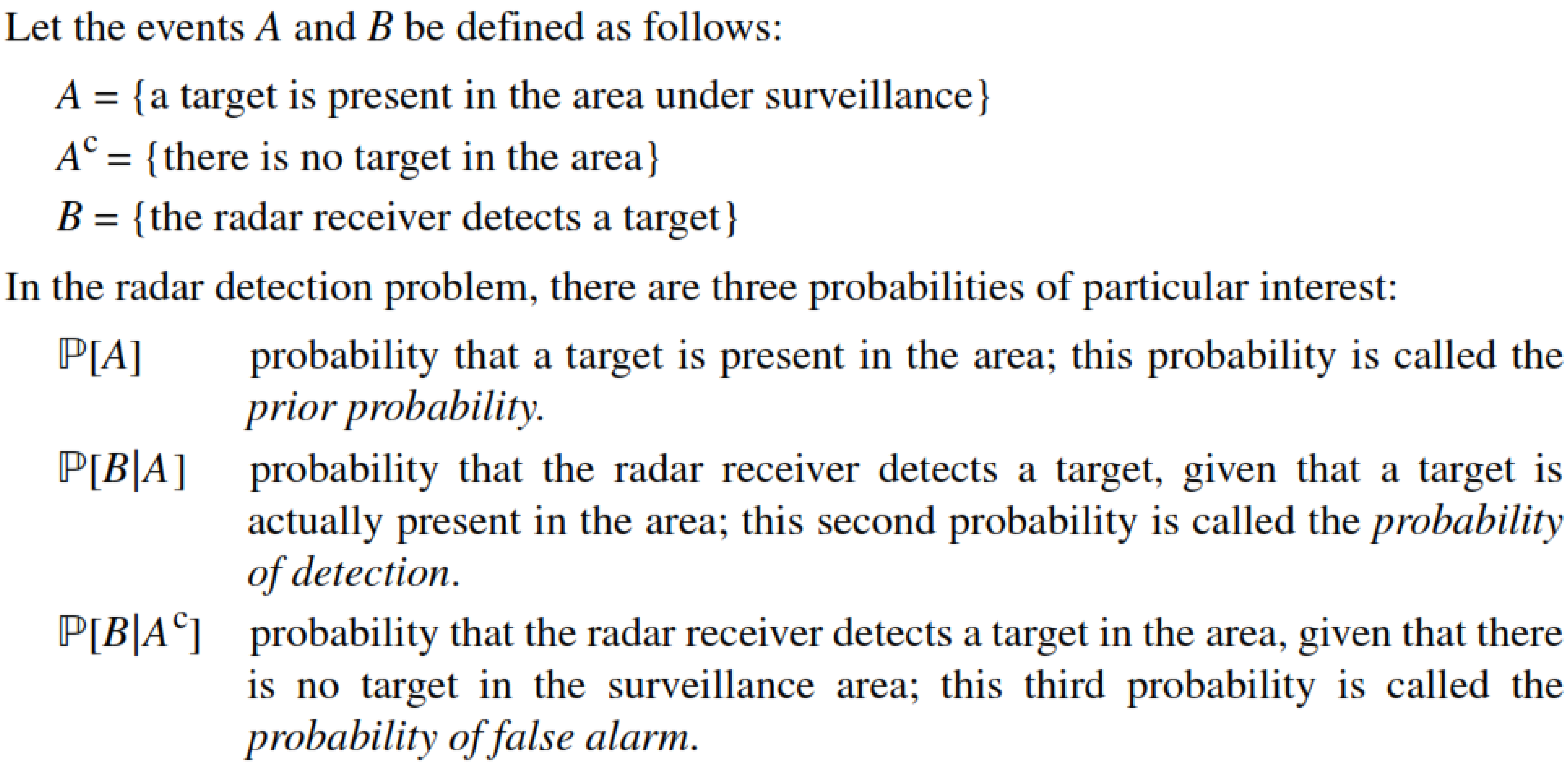
Febraury 26th, 2015

General Notes

- Sections
  - Attend in your section
  - To not loose the section quiz, if you have a hard excuse, before the section, send me a message using the course website “moodal” describing your excuse
  - No transfer is allowed till April 1st.
- Oral Exam
  - Divide yourselves in the section to groups (3-5). Next week is the last week to register a group, otherwise you will be assigned to a group using alphabet order. Give the group names/section to your lecturer assistant
  - In the reference section, include at least one related journal paper with an impact factor larger than 1.5 and note to its impact factor in the presentation

Example of Previous Lecture

Radar, a remote sensing system, operates by transmitting a sequence of pulses and has its receiver listen to echoes produced by a target (e.g., aircraft) that could be present in its surveillance area.



Let the events A and B be defined as follows:

$A = \{$a target is present in the area under surveillance$\}$

$A^{c} = \{$there is no target in the area$\}$

$B = \{$the radar receiver detects a target$\}$

In the radar detection problem, there are three probabilities of particular interest:

$\mathbb{P}[A]$	probability that a target is present in the area; this probability is called the *prior probability.*
$\mathbb{P}[B\|A]$	probability that the radar receiver detects a target, given that a target is actually present in the area; this second probability is called the *probability of detection.*
$\mathbb{P}[B\|A^{c}]$	probability that the radar receiver detects a target in the area, given that there is no target in the surveillance area; this third probability is called the *probability of false alarm.*

Given:

$$\mathbb{P}[A] = 0.02$$
$$\mathbb{P}[B|A] = 0.99$$
$$\mathbb{P}[B|A^{c}] = 0.01$$

Required: $\boxed{\mathbb{P}[A|B]}$

The probability that a target is present in the surveillance area given that the radar receiver has made a target detection.

Question

- Does the relation $P(B|A)+P(B^c|A)=1$?
  - Since conditional probability satisfies the axioms of probability. Then, $P(B|A)+P(B^c|A)=1$
  - Proof: $A=\Omega\cap A=(B\cup B^c)\cap A=(B\cap A)\cup(B^c\cap A)$ **OR** $P(\Omega|A)=P(\Omega\cap A)/P(A)=1$

    $P(A)=P(B|A)P(A)+P(B^c|A)P(A)$

    $1=P(B|A)+P(B^c|A)$

    $P((B^c\cup B)\cap A)/P(A)=1$

    $P(B|A)P(A)+P(B^c|A)P(A)=P(A)$
- Does the relation $P(B|A)+P(B|A^c)=1$?
  - **A contradicting case:** A and B are independent then A^c and B are independent. Why?

    $P(B|A)=P(B|A^c)=P(B)$, if $P(B)\neq 0.5$ then the relation does not hold
  - Note that $P(B|A^c)+P(B^c|A^c)=1$
  - The only condition $P(B|A)+P(B|A^c)=1$ is when $P(B^c|A)=P(B|A^c)$ and $P(B|A)=P(B^c|A^c)$

Lecture Outlines

- Review on probability theory
    - Some important distributions
    - Characteristic function
    - The Gaussian distribution
    - The central limit theorem

Discrete R.Vs Distributions

1. **Binomial:**

- **Parameters:** n (positive integer), p $(0 \leq p \leq 1)$
- **P.m.f.:** $p(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (k = 0, 1, 2, \ldots, n)$
- **Expectation and variance:** $\mu = np$, $\sigma^2 = np(1-p)$
- **Arises as:** Distribution of *number* of successes in success/failure trials ("Bernoulli trials").

e.g. P(2H appears in a series of 5 tosses)=$\binom{5}{2}\, p(\mathrm{H})^2\, p(\mathrm{T})^3$

2. Bernoulli

$$p_X(x) = \begin{cases} 1-p & \text{if } x = 0 \\ p & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \mathbb{E}[X] &= p \\ \operatorname{var}[X] &= p(1-p) \end{aligned}$$

Discrete R.Vs Distributions (cont'd)

3. **Poisson:**

- **Parameter:** $\lambda > 0$
- **P.m.f.:** $p(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (k = 0, 1, 2, \ldots)$
- **Expectation and variance:** $\mu = \lambda$, $\sigma^2 = \lambda$
- **Exponential series formula:** $\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{\lambda}$
- **Arises as:** Distribution of number of occurrences of *rare* events, e.g., accidents, disasters.
- **Poisson approximation to binomial distribution:** A binomial distribution in which n is large and p correspondingly small (so that np is of moderate size) is approximately Poisson with parameter $\lambda = np$; i.e., if X has binomial distribution with parameters n and p as above, then $P(X = k) \approx e^{-\lambda}\lambda^k/k!$, where $\lambda = np$.

Discrete R.Vs Distributions (cont'd)

4. **Geometric:**
    - **Parameter:** p $(0 < p < 1)$
    - **P.m.f.:** $p(k) = (1-p)^{k-1}p$ $(k = 1, 2, \ldots)$
    - **Expectation:** $\mu = 1/p$.
    - **Geometric series formula:** $\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$ $(|r| < 1)$
    - **Arises as:** Distribution of *time (trial) of first success* in a success/failure trial sequence.

Example 1

A binary source generates digits 1 and 0 randomly with probability $P(1)=0.8$ and $P(0)=0.2$

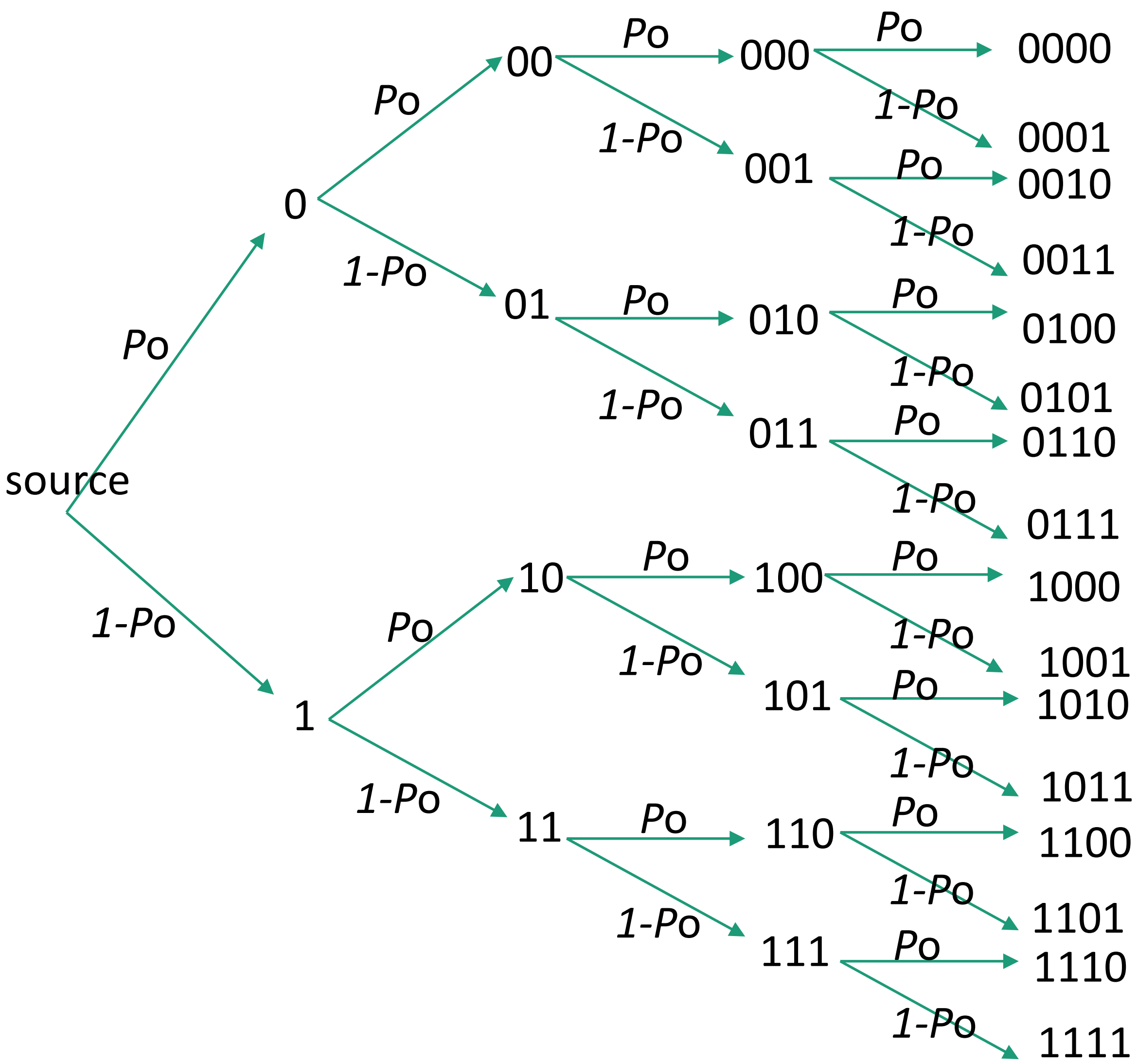
(a) Display graphically the sequential evolution of a four digit sequence. the results

(b) What is the probability that all the four digits are ‘1’s

(c) What is the probability that exactly a ‘1’ will occur in a four digit sequence

(d) What is the probability that at least two ‘1’ will occur in a four digit sequence

Solution



$P(0000)=Po^4$

$P(0001)=Po^3(1-Po)$

$P(0010)=Po^3(1-Po)$

$P(0011)=Po^2(1-Po)^2$

$P(0100)=Po^3(1-Po)$

$P(0101)=Po^2(1-Po)^2$

$P(0110)=Po^2(1-Po)^2$

$P(0111)=Po(1-Po)^3$

$P(1000)=Po^3(1-Po)$

$P(1001)=Po^2(1-Po)^2$

$P(1010)=Po^2(1-Po)^2$

$P(1011)=Po(1-Po)^3$

$P(1100)=Po^2(1-Po)^2$

$P(1101)=Po(1-Po)^3$

$P(1110)=Po(1-Po)^3$

$P(1111)=(1-Po)^4= 0.8^4=0.4096$

Binomial distribution

P(a '1' will occur)

$=P(0001)+P(0010)+P(0100)+P(1000)$

$=4(1-Po)\,Po^3=4\times0.8\times0.2^3=0.0256$

$=\binom{4}{1}(1-Po)\,Po^3$

P(at least two '1's will occur)

=P(two '1's will occur)

+ P(three '1's will occur)

+P(four '1's will occur)

$=P(0011)+P(0101)+P(0110)+P(1001)+P(1010)+P(1100)+$
$+P(0111)+P(1011)+P(1101)+P(1110)$
$+P(1111)$

$=\binom{4}{2}(1-Po)^2\,Po^2+\binom{4}{3}(1-Po)^3Po+\binom{4}{4}(1-Po)^4$

$=6\times0.8^2\times0.2^2+4\times0.8^3\times0.2^3+1\times0.8^4=0.1536+0.0164+0.4096$

$=0.579584$

Example 2 (section problem)

A discrete memoryless channel is used to transmit binary data. The channel is discrete in that it is designed to handle discrete messages and it is memoryless in that at any instant of time the channel depends on the channel input only at that time. Owing to the unavoidable presence of noise in the channel, errors are made in the received binary data stream. The channel is symmetric in that the probability of receiving symbol 1 when symbol 0 is sent is the same as the probability of receiving symbol 0 when symbol 1 is sent.

The transmitter sends 0s across the channel with probability p_o and 1s with probability p_1. The receiver occasionally makes random decision errors with probability p; that is, when symbol 0 is sent across the channel, the receiver makes a decision in favor of symbol 1, and vice versa.

Referring to Figure below, determine the following a posteriori probabilities:

(a) The conditional probability of sending symbol *Ao* given that symbol *Bo* was received.

(b) The conditional probability of sending symbol A_1 given that symbol *B1* was received.

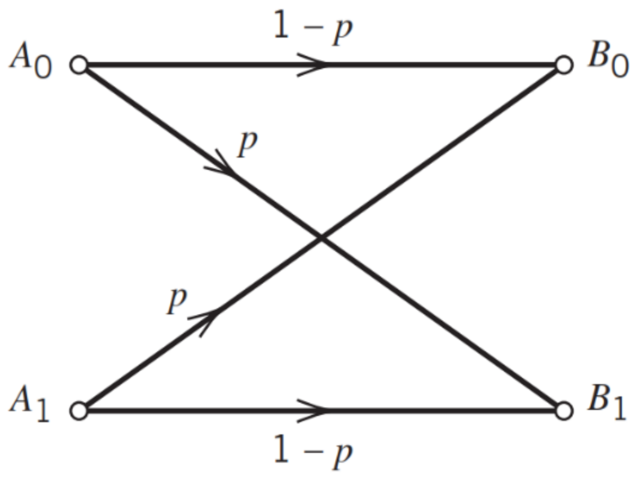
Solution

Assume we have a ***binary symmetric channel*** in that the probability of receiving symbol 1 when symbol 0 is sent is the same as the probability of receiving symbol 0 when symbol 1 is sent. For each bit we have:

Prior probabilities:

P(Ao)=probability that we send the symbol Ao='0' =Po

P(A_1)=probability that we send the symbol A_1='1' =P_1=1-Po



Conditional Probabilities (probabilities of channel error):

P(Bo|A_1)=probability of receiving the symbol Bo='0' given that we sent the symbol A1

P(B_1|Ao)=probability of receiving B1 given that we sent the symbol Ao

Note that from the axiom of probability (normalization probability)

P(Bo|A_1)+P(B_1|A_1)=1

P(Bo|Ao)+P(B_1|Ao)=1

Assumptions (seen in the figure):

Let P(Bo|A_1)=p then P(B_1|A_1)=1-p

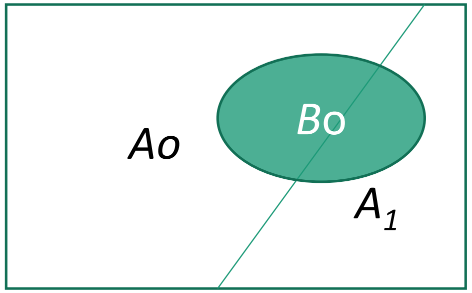
Since the channel is symmetric, P(Bo|Ao)=p thus P(B_1|Ao)=1-p

Solution (cont'd)

Evidence:

P(Bo)=probability of receiving the symbol 'Bo'

P(B_1)=probability of receiving the symbol 'B_1'



P(Bo)=P($Bo|Ao$)P(Ao)+P($Bo|A_1$)P(A_1)=(1-p)Po+p(1-Po)

A Posteriori Probabilities

P($Ao|B_o$)=probability of sending symbol *Ao* given that symbol *Bo* was received

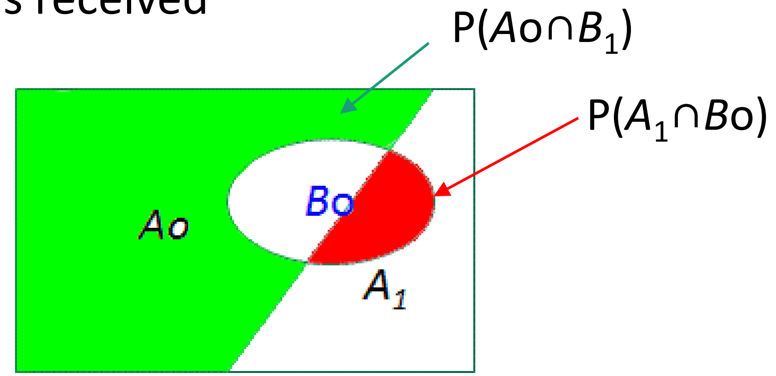
Using Bayes' rule: P($Ao|B_o$)=$P(Bo|Ao)P(Ao)$/P(Bo)

P($A_1|B_1$)=probability of of sending symbol A_1 given that symbol *B1* was received

Channel error=error if a '0' is sent +error if a '1' is sent

P(channel error)= P($Ao \cap B_1$)+P($A_1 \cap Bo$)

=P($B_1|A_o$)P(Ao) + P($Bo|A_1$)P(A_1)



Continuous R.Vs

1. Uniform

$$f_X(x) = \frac{1}{b-a}, \qquad a \le x \le b$$

$$\mathbb{E}[X] = \frac{1}{2}(a+b)$$

$$\mathrm{var}[X] = \frac{1}{12}(b-a)^2$$

2. Exponential

$$f_X(x) = \lambda \exp(-\lambda x), \qquad x \ge 0 \text{ and } \lambda > 0$$

$$\mathbb{E}[X] = 1/\lambda$$

$$\mathrm{var}[X^2] = 1/\lambda^2$$

3. Gaussian

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp[-(x-\mu)^2/2\sigma^2], \qquad -\infty < x < \infty$$

$$\mathbb{E}[X] = \mu$$

$$\mathrm{var}[X] = \sigma^2$$

4. Rayleigh

$$f_X(x) = \frac{x}{\sigma^2}\exp(-x^2/2\sigma^2), \qquad x \ge 0 \text{ and } \sigma > 0$$

$$\mathbb{E}[X] = \sigma\sqrt{\pi/2}$$

$$\mathrm{var}[X] = \left(2 - \frac{\pi}{2}\right)\sigma^2$$

5. Laplacian

$$f_X(x) = \frac{\lambda}{2}\exp(-\lambda|x|), \qquad -\infty < x < \infty \text{ and } \lambda > 0$$

$$\mathbb{E}[X] = 0$$

$$\mathrm{var}[X] = 2/\lambda^2$$

Characteristic function (moment generating function)

- Formulate the probability law of the expectation of the nth moment using pdf:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x \longrightarrow g(X) = X^n \longrightarrow \mathbb{E}[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\,\mathrm{d}x$$

Characteristic function (cont’d)

- Another way is to use the characteristic function:
  - The characteristic function of a continuous random variable *X* is defined as the expectation of the complex exponential function

$$\emptyset_X(v)=\mathrm{E}(e^{-ivx})=\int_{-\infty}^{\infty} f_X(x)e^{ivx}\mathrm{d}x$$

Where *v* is real and j=$\sqrt{-1}$

  - We may also view the characteristic function of the random variable *X* as the Fourier transform of the associated probability density function $f_X(x)$, except for a sign change in the exponent

$$f_X(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} \emptyset_X(v)e^{-ivx}\mathrm{d}v$$

Characteristic function (cont'd)

- $f_X(x)$ and $\Phi_X(\nu)$ forming a **Fourier-transform pair**
- We may obtain the moments of the random variable *X* from the by differentiate both sides of $\Phi_X(\nu)$ with respect to *v* a total of *n* times, and then set *v* = 0; we thus get the result

$$\frac{\mathrm{d}^n}{\mathrm{d}\nu^n}\Phi_X(\nu)\Big|_{\nu=0} = (\mathrm{j})^n \int_{-\infty}^{\infty} x^n f_X(x)\,\mathrm{d}x$$

Moment theorem

$$\boxed{\mathbb{E}[X^n] = (-\mathrm{j})^n \frac{\mathrm{d}^n}{\mathrm{d}\nu^n}\Phi_X(\nu)\Big|_{\nu=0}}$$

Example

- Using the characteristic function, find the mean of the exponential distribution given below:

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x), & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Solution

$$\Phi(\nu) = \int_0^{\infty} \lambda \exp(-\lambda x) \exp(j\nu x)\, dx \longrightarrow \Phi'_X(\nu) = \frac{\lambda j}{(\lambda - j\nu)^2} \longrightarrow \mathbb{E}[X] = -j\Phi'_X(\nu)\Big|_{\nu = 0}$$

$$= \frac{\lambda}{\lambda - j\nu}$$

$$= \frac{1}{\lambda}$$

$$\emptyset_X(v) = \mathrm{E}(e^{-ivx}) = \int_{-\infty}^{\infty} f_X(x) e^{ivx} \mathrm{d}x \qquad f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \emptyset_X(v) e^{-ivx} \mathrm{d}v$$

The Gaussian (Normal) Distribution

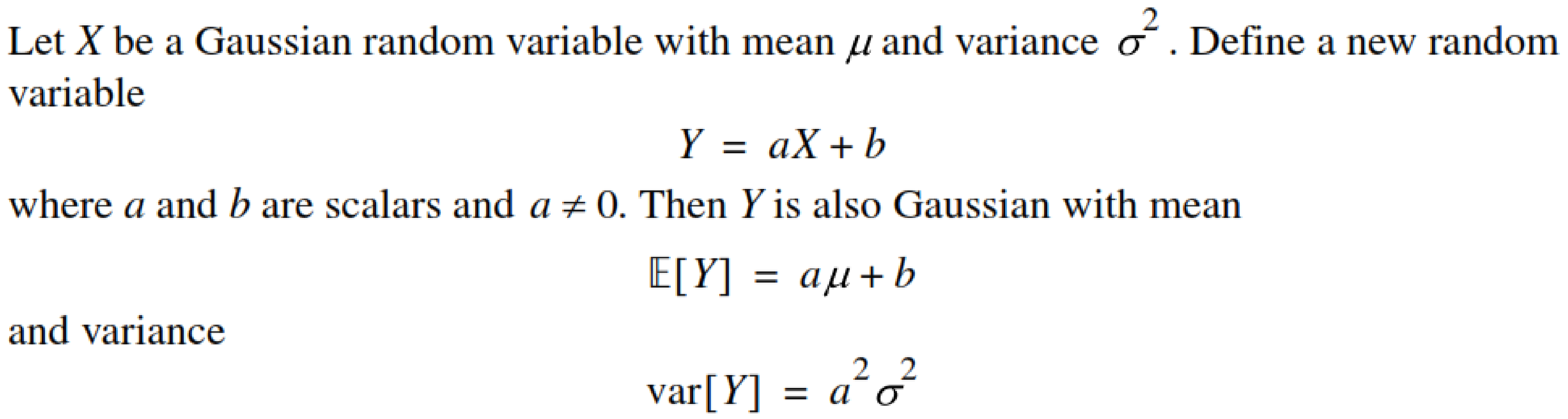
- The **Gaussian distribution** stands out, by far, as the most commonly used distribution in the statistical analysis of communications systems,
- Let *X* denote a continuous random variable; the variable *X* is said to be Gaussian if its pdf has the form:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

  - Where μ is the mean σ is the standard deviation, σ is positive

The Gaussian Distribution Properties

- A Gaussian random variable is uniquely defined by specifying its mean and variance
- Gaussianity is preserved by a linear transformation



Let X be a Gaussian random variable with mean μ and variance σ^2. Define a new random variable

$$Y = aX + b$$

where a and b are scalars and $a \neq 0$. Then Y is also Gaussian with mean

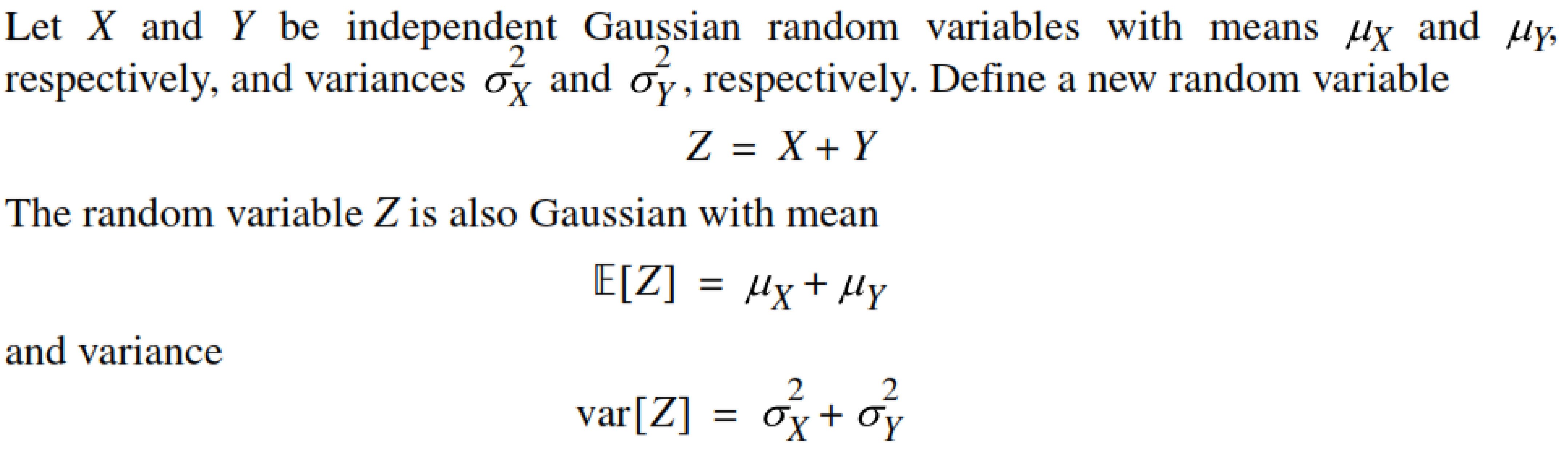
$$\mathbb{E}[Y] = a\mu + b$$

and variance

$$\mathrm{var}[Y] = a^2\sigma^2$$

The Gaussian Distribution Properties (cont’d)

- The sum of independent Gaussian random variables is also a Gaussian random variable, whose mean and variance are respectively equal to the sum of the means and the sum of the variances of the constituent random variables



Let X and Y be independent Gaussian random variables with means μ_X and μ_Y, respectively, and variances σ_X^2 and σ_Y^2, respectively. Define a new random variable

$$Z = X + Y$$

The random variable Z is also Gaussian with mean

$$\mathbb{E}[Z] = \mu_X + \mu_Y$$

and variance

$$\mathrm{var}[Z] = \sigma_X^2 + \sigma_Y^2$$

Jointly Gaussian Random Variables

Let X and Y be a pair of jointly Gaussian random variables with zero means and variances σ_X^2 and σ_Y^2, respectively. The joint probability density function of X and Y is completely determined by σ_X, σ_Y, and ρ, where ρ is the correlation coefficient $\rho(X, Y) = \dfrac{\operatorname{cov}[XY]}{\sigma_X \sigma_Y}$ Specifically, we have

$$f_{X,Y}(x, y) = c\exp(-q(x, y))$$

where the normalization constant c is defined by

$$c = \frac{1}{2\pi\sqrt{1-\rho^2}\,\sigma_X\sigma_Y}$$

and the exponential term is defined by

$$q(x, y) = \frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_X^2} - 2\rho\frac{xy}{\sigma_X\sigma_Y} + \frac{y^2}{\sigma_Y^2}\right)$$

Jointly Gaussian Random Variables (special cases)

- if $\rho=0$: $$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y}\exp\left(-\frac{x^2}{2\sigma_X^2}-\frac{y^2}{2\sigma_Y^2}\right)$$ $$= f_X(x)f_Y(y)$$
  - If the random variables *X* and *Y* are both Gaussian with zero mean and if they are also orthogonal (that is, E[*XY*] = 0), then they are statistically independent

The standard Gaussian Distribution

- When $\mu = 0$ and $\sigma^2 = 1$ $\longrightarrow$ $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ $\longrightarrow$ $F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t$

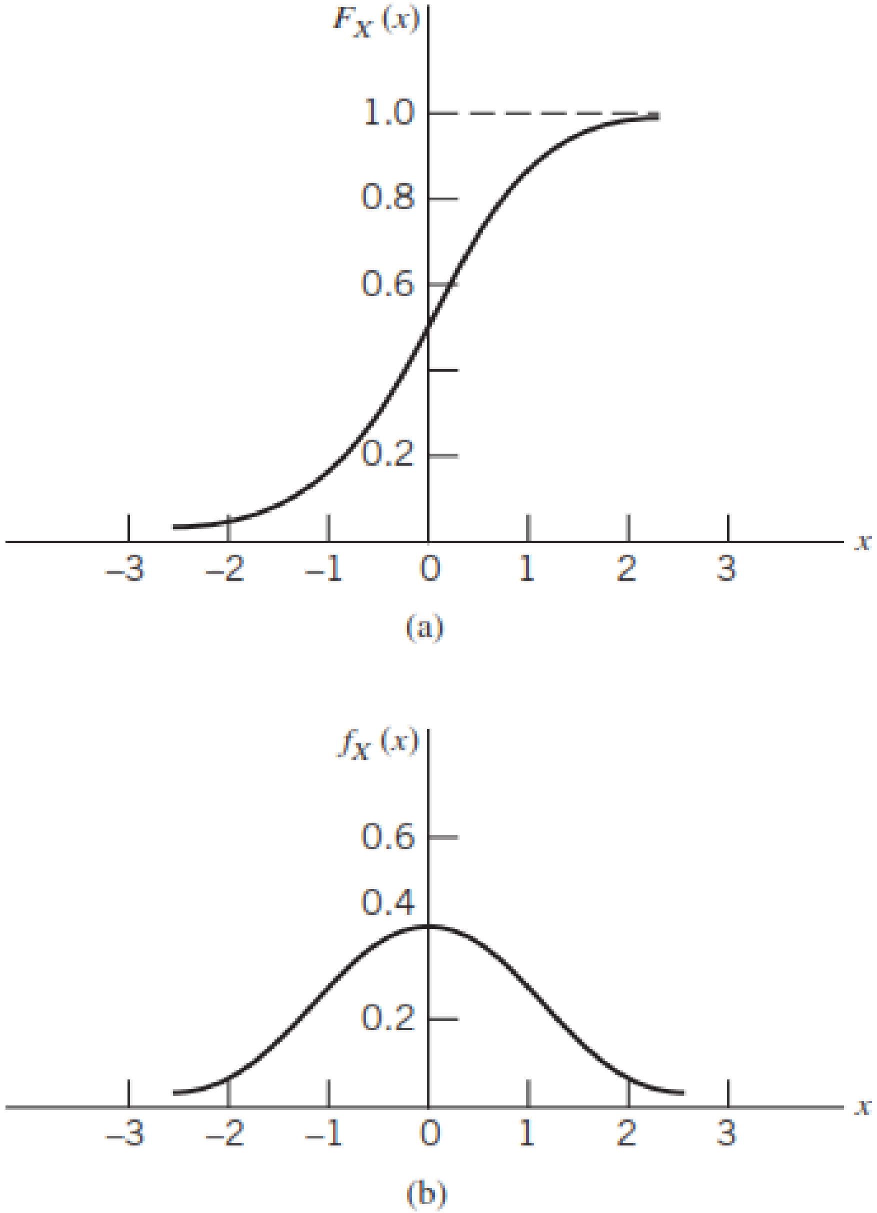
Distribution is defined on its standard form

- The Q-function $Q(x) = 1 - F_X(x)$

$$= \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t$$

- The Q-function, Q(x), is equal to the area covered by the tail of the probability density function of the standard Gaussian random variable X, extending from x to infinity

$$Q(-x) = 1 - Q(x)$$



(a)

(b)

Figure 3.10 The normalized Gaussian (a) distribution function and (b) probability density function.

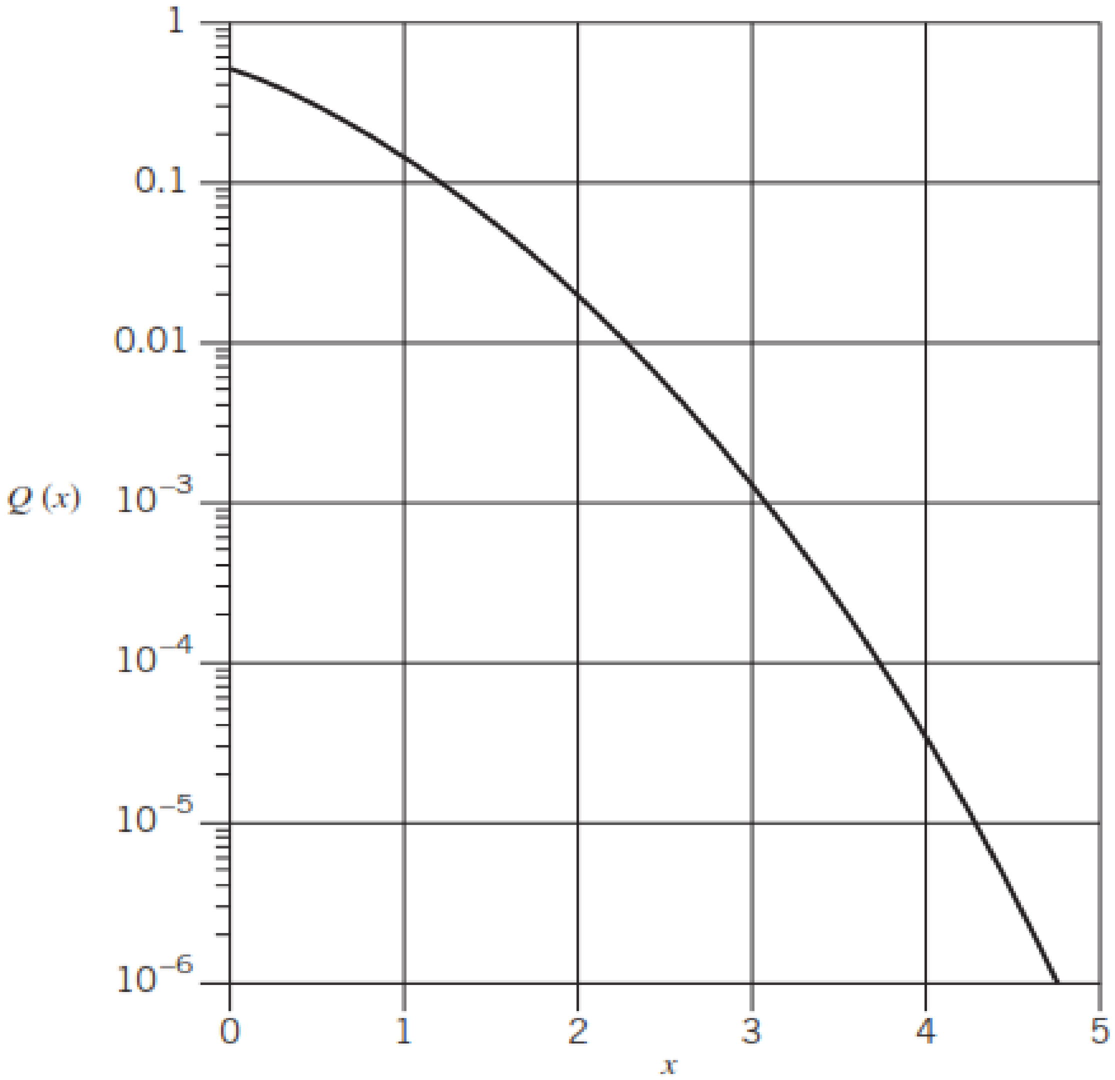


Figure 3.11 The Q-function.

Table 3.1 The standard Gaussian distribution (Q-function) table[5]

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6460	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7485	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9149	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

1. The entries in this table, x say, occupy the range [0.0, 3.49]; the x is sample value of the random variable X.
2. For each value of x, the table provides the corresponding value of the Q-function:

$$Q(x) = 1 - F_X(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} \exp(-t^2/2)\,\mathrm{d}t$$

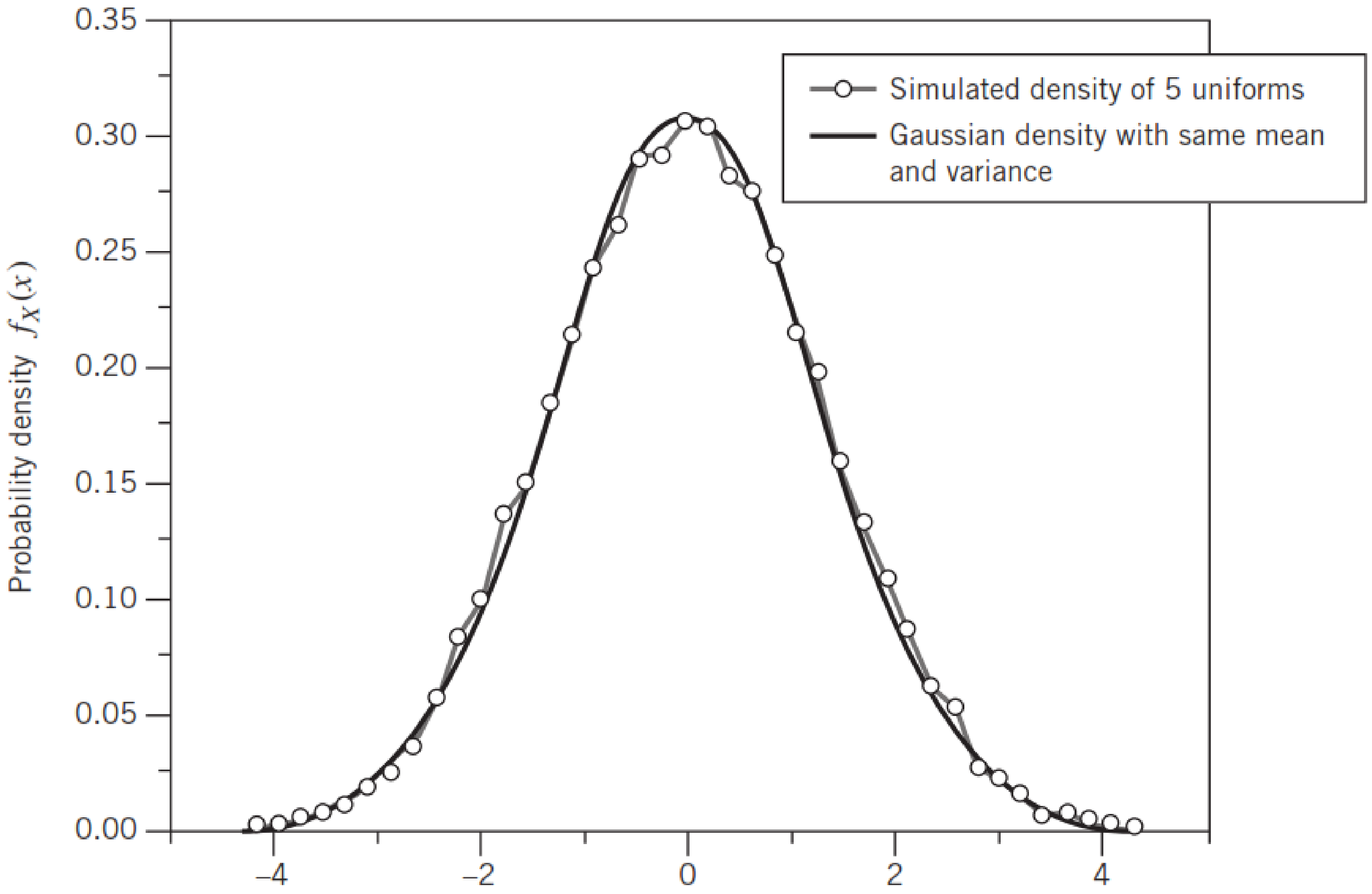
The Central Limit Theorem

- Given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution

- Provides the mathematical justification for using the Gaussian distribution as a model for an observed random variable that is known to be the result of a large number of random events

Example

- Let $Y_n = \sum_{i=1}^{n} X_i$ where the X_i are independent and uniformly distributed random variables on the interval from –1 to +1
- Suppose that we generate 20000 samples of the random variable Y_n for $n = 10$, and then compute the probability density function of Y_n by forming a histogram of the results
- The figure compares the computed histogram (scaled for unit area) with the probability density function of a Gaussian random variable with the same mean and variance
- The figure clearly illustrates that in this particular example the number of independent distributions n does not have to be large for the sum Y_n to closely approximate a Gaussian distribution



Simulation supporting validity of the central limit theorem

Questions